Journal of Biomolecular NMR **29:** 391–394, 2004.
© 2004 *Kluwer Academic Publishers. Printed in the Netherlands.*

Letter to the Editor: Solution structure of the hypothetical protein SAV1595 from *Staphylococcus aureus*, a putative RNA binding protein

Dingjiang Liu* & Daniel F. Wyss
Schering-Plough Research Institute, Department of Structural Chemistry, 2015 Galloping Hill Road, Kenilworth, NJ 07033, U.S.A.

Received 10 December 2003; Accepted 16 January 2004

Key words: NMR structure, RNA binding protein, SAV1595, UPF0044, YhbY

Biological context

The spread of resistant bacterial strains have made conventional antibiotics less effective (Hand, 2000). Therefore, the discovery of new antimicrobial agents has become urgent. Currently, nearly 110 microbial genomic sequences are available as a result of large scale genomic efforts (http://www.ncbi.nlm.nih.gov/PMGifs/Genomes/micr.html, National Center for Biotechnology Information, Bethesda, MD). Analysis of genomic sequences has revealed many evolutionary conserved, functionally uncharacterized protein families which have been cataloged and annotated by several databases such as the Pfam database of protein families (Bateman et al., 2000) and the COG database of clusters of orthologous groups of proteins (Tatusov et al., 2001). Concerted efforts within the fields of bioinformatics, biochemistry, genetics and structural biology have been devoted to understand the structure, function, essentiality and genetics of these protein families. Data derived from these studies may yield potential new drug targets for the development of novel broad-spectrum antibiotics (Freiberg et al., 2001) and are important to the target selection process in the pharmaceutical industry.

SAV1595 is a hypothetical *Staphylococcus aureus* protein. It contains 96 amino acids and has a molecular weight of 11.1 kDa. SAV1595 is a member of a conserved protein family designated as UPF0044 and COG1534 by the Pfam and the COG database, respectively. This protein family is composed of proteins of uncharacterized function with molecular weights ranging from 10 to 15 kDa. Although the members of UPF0044 have been implicated in RNA splicing and translation (Ostheimer et al., 2002), their exact biological functions are still unknown. The crystal structures of two homologs of UPF0044 from *E. coli* (YhbY) and *Haemophilus influenzae* (HI1333) have recently been reported (Ostheimer et al., 2002; Willis et al., 2002). SAV1595 has sequence identities of 38% and 36% with YhbY and HI1333, respectively. YhbY and HI1333 are α/β proteins with an α-β-α-β-α-β-β topology, similar to that of the C-terminal domain of the translation initiation factor 3 which binds RNA. The structural analysis of YhbY and HI1333 indicate that the members of UPF0044 are likely RNA binding proteins. Here, we report the NMR resonance assignments and solution structure of SAV1595. These data will serve as a basis for further structural and functional characterization of the UPF0044 protein family, and for evaluating SAV1595 as a potential target for the screening of customized small molecule libraries by NMR (Shuker et al., 1996).

Methods and results

Recombinant SAV1595 was overexpressed in *E. coli* strain BL21(DE3). The cell cultures were grown at 37 °C in M9 minimal medium with $^{15}NH_4Cl$ as the sole nitrogen source and/or ^{13}C-glucose as the sole carbon source. Proteins were produced and purified as described previously (Liu et al., 2002). The C-terminal His-tag with a sequence of LEHHHHHH was not removed. NMR samples containing 0.3–1.0 mM protein were prepared in 75 mM K_iPO_4, pH 7.5, 5 mM dithiothreitol (DTT), 0.015% NaN_3 and 8% D_2O.

NMR experiments were performed at 25 °C on a Varian *INOVA* 600 MHz spectrometer. The HNCO, CBCA(CO)NH, CBCANH, ^{15}N-edited NOESY-HSQC and ^{15}N-edited TOCSY-HSQC experiments (Proteinpack, Varian Inc.) were used to obtain $^1H^N$, $^1H\alpha$, ^{15}N, $^{13}C'$, $^{13}C\alpha$ and $^{13}C\beta$ resonance assignments. Side chain resonance assignments were obtained based on HCCH-TOCSY, ^{15}N-edited TOCSY-HSQC and $^{13}C/^{15}N$-resolved [^1H,^1H]-NOESY data (Kanelis et al., 2001). The methyl groups of Val and Leu residues were stereospecifically assigned using a

*To whom correspondence should be addressed. E-mail: Dingjiang.liu@spcorp.com

10% ^{13}C-labeled sample (Neri et al., 1989). The NMR data were processed with FELIX980 (Accelrys Inc.) and analyzed with NMRView (Johnson and Blevins, 1994) on a Silicon Graphics workstation.

Backbone resonances for 88 out of 93 assignable non-proline residues were assigned. The assignments for all the side chain NH_2 groups from Asn and Gln residues were obtained. Four peaks in the spectrum were assigned to the LEHH sequence in the C-terminal His-tag. No peaks for residues Leu40, Glu41, Arg43, Glu44, Leu45 were found in the ^{1}H-^{15}N HSQC spectrum. A rather broadened peak was assigned to residue Asn42 based on the ^{13}C chemical shifts from the CBCANH/CBCA(CO)NH experiments. Side chain ^{1}H and ^{13}C resonance assignments are nearly complete for those residues with assigned backbone resonances. Although the ^{1}H-^{15}N cross peaks of Glu41 and Leu45 were absent in the HSQC spectrum, their side chain ^{1}H-^{13}C cross peaks were observed in the ^{1}H-^{13}C HSQC spectrum, and could be assigned based on the ^{13}C chemical shifts and NOE information. The ^{1}H, ^{15}N, and ^{13}C chemical shifts have been deposited in the BioMagResBank (http://www.bmrb.wisc.edu) under accession number 5763.

The solution structure of SAV1595 was determined using a simulated annealing protocol within XPLOR98 (Accelrys Inc.). A ^{15}N-edited NOESY-HSQC spectrum (Figure 1; τ_m = 100 ms, Zhang et al., 1994) and ^{13}C/^{15}N-resolved [^{1}H,^{1}H]-NOESY-HSQC (τ_m = 125 ms, Pascal et al., 1994) spectra in D_2O and H_2O buffer were acquired to obtain the NOE distance restraints. Backbone dihedral angle restraints were obtained using the program TALOS based on the backbone ^{1}Hα, ^{13}Cα, ^{13}Cβ and ^{13}C′ chemical shifts (Cornilescu et al., 1999). The hydrogen bond restraints were determined based on the slow exchanging amide protons identified by recording ^{1}H-^{15}N HSQC two hours after a ^{15}N-labeled SAV1595 protein sample had been exchanged into a D_2O buffer. The distance restraints used for structure calculations included 1338 NOE restraints obtained from an analysis of NOESY datasets and 86 hydrogen bond restraints (Table 1). 10 energy-minimized conformers with the lowest overall energy were selected from 50 calculated structures to represent the solution structure of SAV1595. The overall RMSD values relative to the mean structure of the 10 representative conformers (Figure 2A) were 0.63 ± 0.09 Å and 1.18 ± 0.17 Å, for the backbone atoms (C′, Cα and N) and all heavy atoms, respectively, of the same region. For the residues in the regular secondary structures (residues 4–13(α1), 22–24(β1), 30–42(α2), 45–50(β2), 57–69(α3), 72–77(β3) and 80–84(β4)), the RMSD of backbone and heavy atoms were 0.39 ± 0.05 Å and 0.86 ± 0.05 Å, respectively.

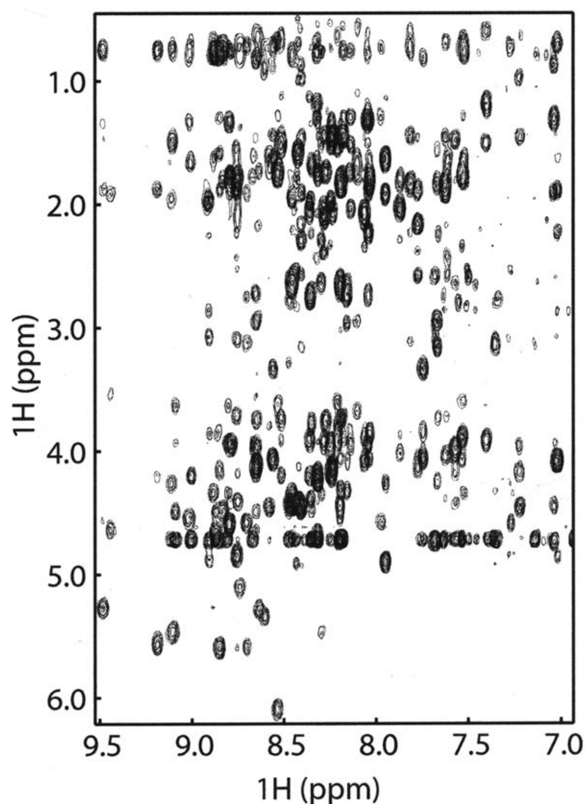

Figure 1. The aliphatic-amide proton region of the projection of the 3D ^{15}N-edited NOESY-HSQC spectrum (τ_m = 100 ms) of SAV1595.

The overall topology of SAV1595 is α-β-α-β-α-β-β. A β sheet is formed by a mixed arrangement of four strands (Figure 2B). Strands 2, 3 and 4 are arranged in antiparallel fashion, while strand 1 is parallel to strand 2. The first helix (residues 4–13) packs against one face of the β sheet, helix 2 (residues 30–42) and helix 3 (residues 57–69) are located on the opposite side. The main interactions of helix 1 and helix 2 with the β sheet are hydrophobic. Residues Leu10 and Leu13 from helix 1 together with Ala14 and Ile17 form a hydrophobic core with the β sheet residues Leu45, Val74, Val82 and Tyr84. The position of helix 1 relative to the β sheet is further determined by additional interactions between Lys7 and Gln75 and Lys11 and Gln75. The β sheet residues Ile23, Ile46, Val48 and Val50 together with residues Ile20 and Phe21 interact with residues Met32, Ile36, Thr39 and Leu40 of

Table 1. Statistics for the NMR structure of SAV1595a

Distance restraints		
Total experimental restraints		1543
All NOE distance restraints		1338
Intraresidue		530
Interresidue sequential ($\lvert i-j \rvert = 1$)		271
Interresidue medium range ($1 < \lvert i-j \rvert < 5$)		195
Interresidue long range ($\lvert i-j \rvert = 5$)		342
Hydrogen bond restraints[b]		86
Dihedral angle restraints		119
ϕ, φ		60,59
CHARMM Lennard-Jones energies (kcal•mol^{-1})[c]		-95 ± 22
Ramachandran analysis (%)		
Residues in favored regions		84.7 ± 1.5
Residues in additional allowed regions		13.7 ± 2.2
Residues in generously allowed regions		1.3 ± 1.2
Residues in disallowed regions		0.9 ± 0.7
Average pairwise r.m.s. deviations (Å)		
	Backbone	All heavy atoms
Residues 1–96	0.63 ± 0.09	1.18 ± 0.17
Residues with secondary structure	0.39 ± 0.05	0.86 ± 0.05

[a]Statistics are based on 10 energy-minimized conformers with the lowest overall energy selected from 50 calculated structures. None of these had violations of distance restraints > 0.5 Å or dihedral angle restraints > 5°.
[b]Two restraints were included for each hydrogen bond.
[c]The CHARMM Lennard-Jones van der Waals energy term, which was not included in the force field of the simulated annealing or restrained minimization, was used to assess the atomic packing in the protein structure.

helix 2 to form another hydrophobic core on the other side of the β sheet. In contrast, helix 3 displays only very limited hydrophobic interactions with the central β sheet through Ala62 with Ile81. The relative orientation of helix 3 is largely determined by the interactions with helix 2. Residues Leu61, Leu65 and Thr69 of helix 3 contact the corresponding hydrophobic surface formed by residues Ile28, Ile33, Leu40 and Val50. It is interesting that the C-terminus of SAV1595 is relatively well structured even in the absence of regular secondary structure. The C-terminal residues are stabilized through both hydrophobic and electrostatic interactions with the rest of the protein. The main hydrophobic interactions involve residues Ile93 and Leu95 which interacts with Leu10 and Tyr84, respectively, while residues Lys88 and Lys91 interact with Glu41 and Ser87. The atomic coordinates have been deposited in the Protein Data Bank (PDB ID code 1RQ8).

Discussion and conclusions

The crystal structures of two other members of the UFP0044 family, namely YhbY and HI1333 of gram-negative bacteria *E. coli* and *H. Influenzae*, respectively, have been solved. SAV1595 is a member of UFP0044 from gram-positive bacteria. The solution structure of SAV1595 and the crystal structures of YhbY and HI1333 are highly similar, superposition of 88 Cα atoms of SAV1595 and YhbY give an rmsd value of 1.60 Å. This is consistent with the high degree of sequence conservation between the members of UFP0044 proteins and confirms the structural conservation between UFP0044 proteins from gram-positive and gram-negative bacteria. As has been discussed for YhbY and HI1333 (Ostheimer et al., 2002; Willis et al., 2002), UFP0044 proteins share a similar fold with several nucleic acid binding proteins, including the C-terminal domain of prokaryotic initiation factor 3, suggesting an RNA binding function for members of the UFP0044 family. Analysis of the electrostatic surface properties revealed a basic molecular surface largely formed by conserved residues Lys7, Arg8, Lys11, Lys47 and Lys58 (Figure 3A). In addition, residues 24 to 27 of SAV1595 with a sequence of GKGG, identified as a putative Gly-X-X-Gly RNA binding motif (Grishin, 2001), resides on the same side of the molecule as the positively charged molecular surface. The Gly-X-X-Gly motif can interact with

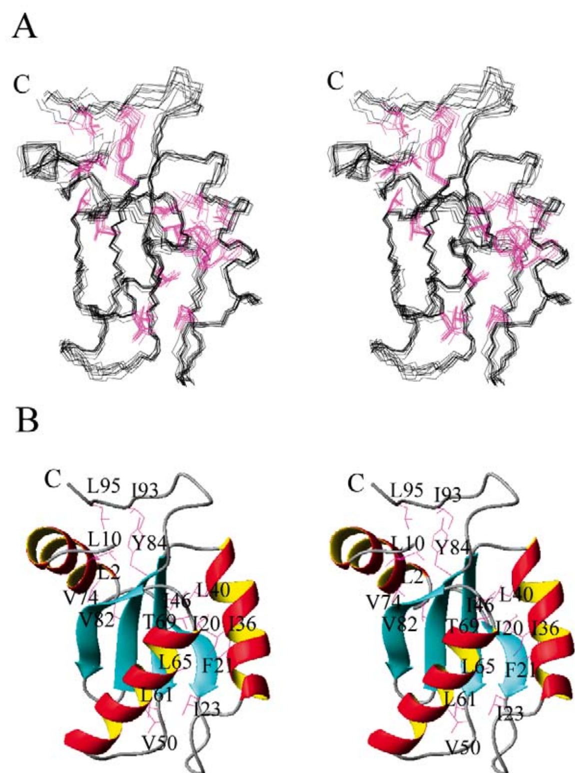

Figure 2. NMR solution structure of SAV1595. (A) Superposition of the ten conformers of SAV1595. (B) A ribbon representation of the NMR structure of SAV1595. The sidechains of selected hydrophobic residues are highlighted in magenta to illustrate the hydrophobic packing. Figures were produced using MOLMOL (Koradi et al., 1996).

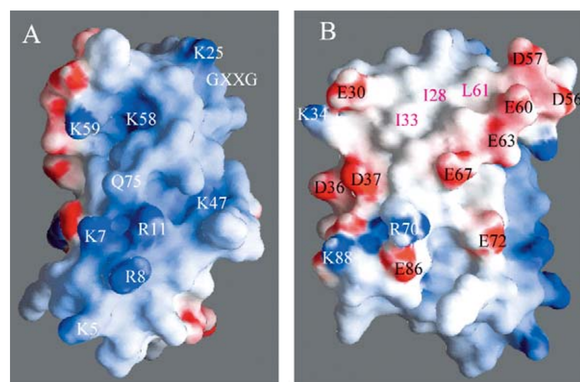

Figure 3. Molecular surface of SAV1595. (A) The putative RNA binding surface involving the positively charged surface areas (blue) and an RNA binding GXXG motif are labeled. The structure is rotated counterclockwise $\sim 180°$ and $\sim 90°$ along the X-axis and Y-axis, respectively, relative to the orientation in Figure 1. (B) The negatively charged molecular surface and hydrophobic patch of SAV1595, the molecule was rotated about 160° anticlockwise along the Z-axis from the orientation in (A). Residues are labeled in red and magenta for negatively charged and hydrophobic surfaces, respectively. Figures were generated using GRASP (Nicholls et al., 1991).

the nucleic acid directly, as observed for the Gly-X-X-Gly motif in Nova KH domains, a sequence specific RNA binding domain (Lewis et al., 2000). The Gly-X-X-Gly motif together with the basic molecular surface form a putative RNA binding surface. Interestingly, a notable acidic surface involving acidic amino acids Asp56, Asp57, Glu60, Glu63 and Glu67 is found next to a small hydrophobic patch, which is formed by Ile28, Ile33 and Leu61 (Figure 3B). Among these residues, Ile28, Ile33, Asp56, Glu60 and Leu61 are conserved within the bacterial members of UFP0044. These molecular surfaces can be used for specific protein-protein interactions.

The three dimensional structure of a given protein target may be useful to suggest the potential function; however, structural data alone is often not sufficient to define its biological function (Laskowski et al., 2003). As indicated by structural analysis, members of UFP0044 contain the necessary structural properties to interact with RNA. However, it remains to be seen if they in fact interact with RNA in a biological environment and if so, with what kind of RNA.

Acknowledgement

We thank Ms Prasanti Repaka for her work on protein purification.

References

Bateman, A. et al. (2000) *Nucl. Acids Res.*, **28**, 263–266.
Cornilescu, G. et al. (1999) *J. Biomol. NMR*, **13**, 289–302.
Freiberg, C. et al. (2001) *J. Mol. Microbiol. Biotechnol.*, **3**, 483–489.
Grishin, N.V. (2001) *Nucl. Acids. Res.*, **29**, 638–643.
Hand, W.L. (2000) *Adolesc. Med.*, **11**, 427–438.
Johnson, B.A. and Blevins, R.A. (1994) *J. Biomol. NMR*, **4**, 603–614.
Kanelis, V. et al. (2001) *IUBMB Life*, **52**, 291–302.
Koradi, R. et al. (1996) *J. Mol. Graph.*, **14**, 51–55.
Laskowski, R.A. et al. (2003) *J. Struct. Funct. Genomics*, **4**, 167–177.
Lewis, H.A. et al. (2000) *Cell*, 100, 323–332.
Liu, D. et al. (2002) *J. Biomol. NMR*, **23**, 159–160.
Neri, D. et al. (1989) *Biochemistry*, **28**, 7510–7516.
Nicholls, A. et al. (1991) *Proteins*, **11**, 281–296.
Ostheimer, G.J. et al. (2002) *Structure*, **10**, 1593–1601.
Pascal, S.M. et al. (1994) *J. Magn. Reson.*, **B103**, 179–201.
Shuker, S.B. et al. (1996) *Science*, **274**, 1531–1534.
Tatusov, R.L. et al. (2001) *Nucl. Acids Res.*, **29**, 22–28.
Willis, M.A. et al. (2002) *Proteins: Structure, Function, and Genetics*, **49**, 423–426.
Zhang, O. et al. (1994) *J. Biomol. NMR*, **4**, 845–858.